

4

AD-A197 653

INTEGRATING SYNTAX, SEMANTICS, AND DISCOURSE  
DARPA NATURAL LANGUAGE UNDERSTANDING PROGRAM

R&D STATUS REPORT  
Unisys/Defense Systems

ARPA ORDER NUMBER: 5262

PROGRAM CODE NO. NR 049-802 dated 10 August 1984 (433)

CONTRACTOR: Unisys Defense Systems

CONTRACT AMOUNT: \$1,192,833

CONTRACT NO: N00014-85-C-0012

EFFECTIVE DATE OF CONTRACT: 4/29/85

EXPIRATION DATE OF CONTRACT: 4/28/89

PRINCIPAL INVESTIGATOR: Dr. Lynette Hirschman

PHONE NO. (215) 848-7554

SHORT TITLE OF WORK: DARPA Natural Language Understanding Program

REPORTING PERIOD: - 2/1/88 - 4/30/88

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

DTIC  
ELECTE  
JUN 21 1988  
S D

DISTRIBUTION STATEMENT A  
Approved for public release  
Distribution Unlimited

## 1. Description of Progress

We are continuing to port Pundit to the Trident domain of maintenance reports while at the same time extending coverage in the RAINFORMS and CASREPS domains. Parallel development in the three domains has had a synergistic effect on the overall system, as features implemented for one domain become available for the others.

### 1.1. Grammar

An extension of the ISR to cover infinitival complements of adjectives has been implemented and is currently being tested. This will allow us to interpret phrases like 'unable to turn'.

A more general syntactic treatment of object gaps has been implemented. It is not significantly less efficient than our previous lexical treatment. Object gaps are interpreted by treating them as pronouns.

### 1.2. Syntax/Semantics Interaction

A design document for the proposed semantically based selection scheme was written. Implementation is in progress.

Modifications were made to Pundit so that the labeling of subjects and objects is done directly in the ISR rather than being added later by an ad hoc procedure prior to semantic interpretation.

A testing procedure for the selection mechanism has been implemented to uncover gaps in coverage by batch parsing of sentences without selection and xor. Revisions to the selection mechanism have been designed to make it nearly deterministic, which should result in a significant speedup when the revisions are implemented.

We have implemented a component which provides an association of quantifiers and the variables they bind which preserves the order of the quantifiers. This is in preparation for the general treatment of quantifiers which is planned for the third quarter.

### 1.3. Semantics

Noun phrase semantics was extended to allow the dates in TFR headers to receive a semantic analysis.

A capability has been added to PUNDIT for processing isolated noun phrases as responses to prompts, such as *Cause of failure: a bad head assembly wire* (Trident), without a commitment to an existential analysis. This capability is also used for processing noun phrases occurring in message headers.

The part hierarchy for the Trident MTU equipment was entered into the domain model. A semantic treatment of some types of appositive noun phrases, such as *Interlock switch (SW 7802)*, has been implemented.

Semantics rules for 8 new messages were added to the RAINFORM message domain, thus doubling our coverage in that domain.

Rules for three new CASREPS have been added to the CASREPS domain. The semantic treatment of left modifiers has been extended so that adjectives can be interpreted by semantic rules. This included revisions to several problematic aspects of the semantic interpretation of left modifiers.

### 1.4. Time

The manipulation of the data structure for representing the current discourse context which occurs prior to each call to semantic analysis has been changed so that focusing information and temporal information are handled independently.

The representation of aspect in processes has been changed to be independent of the argument structure by introducing a 'doP' operator.

### 1.5. Discourse

The semantic interpreter was enhanced to update the discourse context after processing each clause in a co-ordinate structure. The change increases Pundit's ability to interpret run-on sentences, which are analyzed as paratactic co-ordination. In addition, it provides a more correct treatment of focusing in ordinary conjunction.

Entities in the header of TFR messages are now represented as being in global focus, which makes them available for pronominal reference and as implicit thematic role fillers within the message body. Locally focused entities, in contrast, are available only within their particular discourse segment. Domain-specific procedures have been implemented to order the global foci by saliency in the domain.

We have implemented a focus order interpreter in order to support a declarative representation of the order of expected foci in the focusing algorithm. This has made it much easier to detect and correct several problems with the focus order.

The user interface developed for processing Trident TFRs has been ported to the RAINFORM message domain.

### 1.6. Demo Environment

A graphical display capability for the temporal component has been implemented. It displays temporal relationships among situations mentioned in texts. The graphical tree display in the demo interface has been resized so that much more of the tree is visible on the screen.

### 1.7. Extensions Funded by Other Sources

As part of our 1988 IR&D work, we have added a new meta-rule capability to handle wh-constructions (questions, indirect questions, relative clauses) which is integrated with conjunction. This provides a large increase in the grammar's coverage, and gives us a basis for realistic handling of interactive applications, which normally contain a large proportion of questions.

## 2. Change in Key Personnel

None.

## 3. Summary of Substantive Information from Meetings and Conferences

### 3.1. Darpa Meetings

Shirley Steele, Martha Palmer, and Lynette Hirschman attended the meeting of the Darpa Natural Language contractors at Mohonk, New York, May 4-6.

### 3.2. Papers and Presentations

- (1) Dahl, D. "Natural Language Processing for Database Generation: The PUNDIT System", Presented at AI West, May, 1988, Long Beach California.
- (2) Hirschman, L. "A Meta-Treatment of wh-Constructions". Submitted to META 88 Workshop on Meta Programming in Logic Programming.
- (3) Hirschman, L., Hopkins, W.C., Smith, R.C. "Or-Parallel Speed-up in Natural Language Processing: A Case Study". To be presented at the 5th International Logic Programming Conference, Seattle, August, 1988.

- (4) Lang, F., and Hirschman, L. "Improved Portability and Parsing Through Interactive Acquisition of Semantic Information". Presented at the Second Conference on Applied Natural Language Processing, Austin, Texas, February, 1988.
- (5) Palmer, M., and Linebarger, M. "Status of Verb Representations in PUNDIT". Presented at Theoretical And Computational Issues in Lexical Semantics, Brandeis University, Waltham, Mass, April 21-24, 1988.

### 3.3. Conference Attendance

Shirley Steele attended the International Conference on Acoustics, Speech and Signal Processing in New York in April.

Francois Lang and Martha Palmer attended the Applied ACL meeting in February in Austin. PUNDIT was demonstrated at this conference.

Deborah Dahl attended the Fourth Artificial Intelligence and Advanced Computer Technology Conference, Long Beach, California, May 4-6.

### 4. Problems Expected or Anticipated

None.

### 5. Action Required by the Government

None.

### 6. Fiscal Status

(1) Amount currently provided on contract:  
\$ 1,192,833 (funded)

\$1,704,901 (contract value)

(2) Expenditures and commitments to date:  
\$ \$ 1,045,216

(3) Funds required to complete work:  
\$ 659,685

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By <i>per ltr</i>	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

